Sample Identifiers and Metadata to Support Data Management and Reuse in Multidisciplinary Ecosystem Sciences

Joan E. Damerow[1], Charuleka Varadharajan[1], Kristin Boye[2], Eoin L. Brodie[1], Madison Burrus[1], K. Dana Chadwick[1,3], Robert Crystal-Ornelas[1], Hesham Elbashandy[4], Ricardo J. Eloy Alves[1], Kim S. Ely[5], Amy E. Goldman[6], Ted Haberman[7], Valerie Hendrix[4], Zarine Kakalia[1], Kenneth M. Kemner[8], Annie B. Kersting[9], Nancy Merino[9], Fianna O'Brien[4], Zach Perzan[3], Emily Robles[1], Patrick Sorensen[1], James C. Stegen[6], Ramona L. Walls[10], Pamela Weisenhorn[8], Mavrik Zavarin[9], and Deborah Agarwal[4]

[1]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA;
[2]SLAC National Accelerator Laboratory, Menlo Park, CA;
[3]Department of Earth System Science, Stanford University, Palo Alto, CA;
[4]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA;
[5]Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY;
[6]Pacific Northwest National Laboratory, Richland, WA;
[7]Metadata Game Changers, Boulder, CO;
[8]Molecular Environmental Science, Argonne National Laboratory, Lemont, IL;
[9]Lawrence Livermore National Laboratory, Livermore, CA;
[10]Bio5 Institute, University of Arizona, Tucson, AZ

Contact: Joan Damerow (JoanDamerow@lbl.gov)

Abstract

Physical samples are foundational entities for research across biological, Earth, and environmental sciences. Data generated from sample-based analyses are not only the basis of individual studies, but can also be integrated with other data to answer new and broader-scale questions. Ecosystem studies increasingly rely on multidisciplinary team-science to study climate and environmental changes. While there are widely adopted conventions within certain domains to describe sample data, these have gaps when applied in a multidisciplinary context. In this study, we reviewed existing practices for identifying, characterizing, and linking related environmental samples. We then tested practicalities of assigning persistent identifiers to samples, with standardized metadata, in a pilot field test involving eight United States Department of Energy projects. Participants collected a variety of sample types, with analyses conducted across multiple facilities. We address terminology gaps for multidisciplinary research and make recommendations for assigning identifiers and metadata that supports sample tracking, integration, and reuse. Our goal is to provide a practical approach to sample management, geared towards ecosystem scientists who contribute and reuse sample data.

Introduction

The study of natural ecosystems requires multidisciplinary science teams to understand and model processes from molecular to global scales (Weart, 2013). Many research activities involve diverse collections of samples and associated field or laboratory measurements (Devaraju et al, 2016; Ponsero et al, 2020). For example, studies of organic matter cycling through plants and soil involves analysis of samples to represent soil biogeochemistry, microbial communities, plant structures, leaf gas exchange, and traits of the specific organisms involved (Cordeiro et al, 2020; Malik et al, 2020; Treseder et al, 2012). Each scientific expert, project team, and discipline has a responsibility to ensure that others can interpret, integrate, and reuse their sample data to help solve emerging problems as our global environment continues to change (Soranno and Schimel, 2014).

Collaboration across disciplines requires a more unified approach to report basic information about key data entities, such as samples. One challenge in promoting a unified way of reporting sample data is that some research communities have already developed community-specific conventions, including those for ‘omics samples (Field et al, 2011; Reddy et al, 2015; Yilmaz et al, 2011), biodiversity records (Wieczorek et al, 2012), and geoscience samples (Devaraju et al, 2016; System for Earth Sample Registration (SESAR), 2020a). A larger challenge is that many researchers use no formal reporting conventions, or exclude information needed to interpret and reuse the data (Roche et al, 2015). More coordination is needed across these communities to develop a multidisciplinary reporting format for physical samples that is widely adopted, or to ensure that standards are interoperable. Common reporting would support effective discovery, integration, and reuse of sample data that spans scientific domains.

Sample identifiers are also needed to associate and manage important information describing a sample (i.e. metadata), such as the location, date, environmental context, and purpose of sample collection. For multidisciplinary studies, the task of generating and managing

unique sample identifiers and associated metadata can be complicated, particularly as important contextual information is added throughout the data lifecycle (Treloar and Klump 2019). Samples are sent to different collaborators, laboratories and user facilities, and then combined into a variety of digital records and publications (Figure 1; Chase et al, 2016). As a result, scientists face challenges with (meta)data management, tracking, or the ability to integrate and reuse valuable sample data. Without attention, these inefficiencies result in (meta)data loss and inhibit the potential of scientific discovery.

Our overall goal was to address sample identification and metadata needs of ecosystem scientists, and was driven by the user community of the US Department of Energy's (DOE's) data repository for earth and environmental sciences—Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE; Varadharajan et al, 2018). The DOE's Environmental Systems Science (ESS) program relies on multidisciplinary, team-based science to study complex processes within terrestrial ecosystems, spanning from the bedrock through the rhizosphere and vegetation to the atmospheric surface layer (Biological and Environmental Research Advisory Committee (BERAC), 2017). This community is well-positioned to help address specific challenges in standardizing and integrating (meta)data about a variety of environmental samples (e.g. soil, water, plant, and associated biological material used for 'omics analyses), which applies broadly to environmental research (Chadwick et al, 2020; Serbin et al, 2019; Stegen and Goldman, 2018; Wu et al, 2020, 2019).

We focus on sample identifiers and metadata that support findability, accessibility, interoperability, and reusability (FAIR) from the multidisciplinary domain-science perspective (Beck et al, 2020; Conze et al, 2017; Lehnert et al, 2019b; Stall et al, 2019; Wilkinson et al, 2016). We therefore use a community-focused approach to: a.) evaluate existing options for sample identifiers and metadata descriptions for ecosystem science samples; b.) pilot the process of standardizing sample information to evaluate practical issues from domain-science perspectives; and c.) outline practical recommendations for sample identifier allocation, tracking, and associated metadata.

Methods

Review of existing sample identifiers, metadata conventions and standards

ESS-DIVE's work on sample identifiers and metadata began in response to a specific problem with tracking multidisciplinary samples as they are sent to different labs and user facilities, which DOE ESS scientists brought up during community meetings. As a community-focused data repository, our approach to this issue involved leading or participating in a variety of community discussions on sample identifiers and/or associated metadata. These included: presenting identifier options in an ESS community webinar and whitepaper, discussion with each pilot test participant, several meetings with US DOE user facilities and data systems representatives (Joint Genome Institute, National Microbiome Data Collaborative, Environmental Molecular Sciences Laboratory, and DOE Systems Biology Knowledgebase), broader community meetings on identifier and metadata practices for physical samples [Earth Science Information Partners (ESIP), and Research Data Alliance (RDA)], National Microbiome Data Collaborative (NMDC) Ontology workshop, USGS workshop on sample collection

metadata for the National Digital Catalogue, and participation in the IGSN 2040 Steering Committee and business planning.

After reviewing the scope and use of available identifier options (Table 1) and community discussions, we focused additional identifier comparison on International GeoSample Numbers (IGSNs) and Archival Resource Keys (ARKs), which are most commonly used for a variety of sample types (Supplemental Table 1). Considerations in the identifier assessment included: i.) association with a broader international community focused on sample identification and description, ii.) associated metadata to describe samples and their relationships, iii.) availability of user-friendly infrastructure to mint identifiers and validate metadata, iv.) general ease of use, and v.) other technical identifier characteristics listed in Supplemental Table 1.

We also reviewed existing metadata standards and templates that are relevant for samples collected by environmental scientists, including: general digital object standards (DataCite Metadata Working Group, 2019; DCMI Usage Board, 2020; Open Geospatial Consortium Inc., 2010), biodiversity records (Darwin Core Task Group, 2014; Wieczorek et al, 2012), 'omics (e.g. genomics, metagenomics) material (Field et al, 2011; Reddy et al, 2015; Yilmaz et al, 2011), and geoscience samples (System for Earth Sample Registration (SESAR), 2020a, 2020b) (Supplemental Table 2). We created a translation table comparing 49 metadata elements (Supplemental Table 3) in human-readable format. The translation table depicts linkages where metadata elements were common across standards, and differences.

The core IGSN Descriptive Metadata Schema (https://github.com/IGSN/metadata) includes basic metadata associated with sample collection, which is generally relevant across sample types. This schema links metadata profiles that differ across six currently-functioning IGSN allocating agents. SESAR (the first allocating agent) has no access restrictions for obtaining IGSNs and provides user-friendly services for sample management (https://www.geosamples.org/). The SESAR metadata profile and controlled terms are currently focused on geoscience samples, but the IGSN organization seeks to accommodate multiple disciplines and has already expanded into plant and soil samples for some IGSN allocating agents. Our translation table for sample metadata allowed us to identify metadata elements and terms that could be revised or extended within the SESAR profile for improved representation of other sample types (Supplemental Table 3).

Biology-related standards are well-established, commonly used in the community, and are particularly important for ecosystem science samples. Genomic and metagenomic analyses and data publication require use of standards developed by the Genetic Standards Consortium (GSC) (Field et al, 2011), namely Minimum Information about any Sequence (MIxS) and Minimum Information about any Metagenome (MIMS) (Yilmaz et al, 2011). DarwinCore is a metadata standard for biodiversity records that has been widely adopted across the biocollections community (Wieczorek et al, 2012). It is also required for submitting data to the Global Biodiversity Information Facility (GBIF, www.gbif.org), which allows global search and integration of biodiversity records (Gaiji et al, 2013; Robertson et al, 2014). GBIF provides a valuable service as a data aggregator, and thus has driven standards adoption, and enabled a wide range of data reuse applications in published biodiversity studies (Ball-Damerow et al, 2019; Gaiji et al, 2013), including over 5,000 known citations from studies using biodiversity records (www.gbif.org).

We researched ontologies that could be used to describe a broad set of environmental sample types, including the Biological Collections Ontology (BCO) (Walls et al, 2014),

Environment Ontology (ENVO) (Buttigieg et al, 2016), Population and Community Ontology (PCO; http://purl.obolibrary.org/obo/pco.owl), and Plant Ontology (PO) (Avraham et al, 2008) to identify additional or alternate terms to generally describe other types of soil, sediment, water, gas, and biology-related samples (Damerow et al, 2020).

We also engaged with the broader, international community working on sample-related standards. This broader community is led by members of the IGSN organization, with participation across other national agencies (e.g. USGS, CSIRO, Australia Research Data Commons-ARDC) and data organizations (ESIP and RDA). This community participation was important in identifying best practices in identifier and metadata use, and contributing perspectives of ecosystem sciences in the broader community working on sample standardization. Continued participation in the broader informatics and domain science communities is important for improving interoperability and usability of sample-related standards.

Sample identifier and metadata testing in the field

In order to develop a sample metadata reporting format that was informed by our domain science community, we worked with scientists from eight different Environmental Systems Science projects to conduct a pilot test for using sample PIDs and metadata. In particular, we tested the practicality of the IGSN, which appeared to be the best choice amongst relevant PIDs for our purposes. These projects had varying scopes and sample types, and were all funded by DOE's Office of Science Environmental Systems Science (ESS) program (Supplemental Table 4).

Prior to sample registration, we discussed the following with representatives from each project: 1) expected sample types involved, 2) how to assign IGSNs and link related samples, 3) essential metadata needed to understand specific sample types, and 4) past sample tracking workflows. Some projects had already collected samples and preferred to register for IGSNs after collection to be associated with digital files, while other projects pre-registered their samples before collection, or registered directly after collection. We used initial feedback and background research to identify several core descriptive sample metadata fields likely to be necessary for searches on ESS-DIVE to be most effective, including standardized information on the following (Damerow et al. 2020, and see Supplemental Table 3 for full translation table comparing metadata elements from existing standards and templates):

- IGSN and Parent IGSN (where relevant)
- Sample Name (project-specific sample name, must be unique)
- Chief Scientist/Collector
- Sample Type fields:
    - Object Type (e.g. Individual sample, core, site),
    - Material (e.g. Liquid-aqueous, Rock, Soil, Biology),
    - Sampled Feature (primary physiographic feature sample collected from)
- Location Information (Latitude, Longitude in WGS84; Location description),
- Date (ISO 8601; e.g. 1954-04-07),
- Collection Method Description
- Project

Note that this list represents the initial IGSN metadata fields that should be required, and were subsequently revised after our pilot test work. Many additional metadata fields are available and are recommended or optional depending on the sample type (System for Earth Sample Registration (SESAR), 2020a).

The researchers involved in our testing used SESAR's sample management portal (MySESAR, http://www.geosamples.org/mysesar) to register samples and update metadata. We recommended a specific workflow for participants to register their samples and update sample collection metadata, outlined in our github repository (https://github.com/ess-dive-community/essdive-sample-id-metadata) and associated dataset (Damerow et al. 2020).

We also worked with individuals to map sample history from collection of samples in the field through a variety of analyses, and publication (Figure 2). This exercise helped determine sample tracking needs, and develop recommendations for assigning PIDs and linking highly-related samples and subsamples.

After sample collection and registration, we discussed the following: 1) What sample collection metadata is needed to understand resulting sample data?; 2) How much effort did it take to register samples and standardize metadata?; 3) What is needed to make sample PID registration and standardization easier?

Developing the final IGSN-ESS reporting guidelines

We used a combination of research on existing standards, and pilot test feedback to develop final recommendations for allocating identifiers and assigning standard metadata, summarized in Figure 5 (Damerow et al. 2020). We took extensive notes during meetings with pilot test participants, and compiled specific feedback on improving guidance on allocating identifiers and relationships, metadata needed to understand relevant sample types, and improve efficiency of sample registration and standardization. Pilot test participants identified metadata elements that needed to be added, modified, or removed to improve relevance for multidisciplinary ecosystem science samples. We then used our translation table (Supplemental Table 3) comparing other existing standards to guide specific recommendations. For example, to address feedback regarding inefficiencies in providing all metadata at individual sample levels, we added the Darwin Core elements: Location ID, Collection ID, and Event ID. We then reviewed existing, commonly-used ontologies (ENVO, BCO, PO) to select important vocabulary terms to characterize sample type, material, and environmental context. We developed a list of relevant terms based on pilot test studies, and all participants helped decide on our final term lists for object type and material, specifically.

All feedback was addressed in our final recommendations, which we compiled into github, and more user-friendly gitbook documentation. This documentation includes: instructions on registering samples for IGSNs using our revised template, specific definitions/instructions/examples for each metadata element, lists of terms for elements where controlled vocabulary is needed, and instructions for how to contribute feedback using github, and cite the final format. To develop documentation, we used the ESS-DIVE community github for samples, inspired from user-friendly documentation for Darwin Core, which facilitates additional community feedback (through public github issues) and versioning. We presented our final recommendations and documentation in two additional community webinars, which are advertised to ESS-DIVE users and ESS scientists, and published on the ESS-DIVE website

(https://ess-dive.lbl.gov/webinars/). The purpose of community webinars was to present our conclusions and collect any additional feedback.

As a community-oriented data repository, we will continue to gather feedback and develop additional tools to support users in submitting, searching for, integrating, and reusing high-quality sample data.

Results

Review of Existing Sample Identifier and Metadata Practices

In our review, we found that numerous studies have documented that persistent identifiers (PIDs) enable sample tracking across facilities and publications, and support reuse over time (Conze et al, 2017; Devaraju et al, 2017, 2016; Duerr et al, 2011; Guralnick et al, 2014, 2015; Lehnert et al, 2019a; McMurry et al, 2017; Michener, 2015). PIDs are globally unique, stored with descriptive metadata, and arguably essential for supporting data synthesis (Guralnick et al, 2014; Lehnert et al, 2019a). While there are several options for obtaining PIDs—Archival Resource Keys (ARKs), Digital Object Identifiers (DOI), Uniform Resource Identifier (URI) (Guralnick et al, 2014, 2015; Klump et al, 2016; Lehnert et al, 2019a; McMurry et al, 2017)—the International GeoSample Number (IGSN) is the primary PID for physical samples (Ferguson et al, 2018; Goldstein et al, 2014; Table 1, Supplemental Table 1; Lehnert et al, 2019a). IGSNs were originally designed for geoscience samples, but have been used for a variety of biological and environmental sample types. The IGSN organization is now expanding to better support multidisciplinary samples, and leading the Internet of Samples project (Walls et al. 2020).

Through community discussions, we determined that the most important factors in selecting a PID were a.) an international community with expertise on sample documentation, b.) associated sample-specific metadata that will eventually enable global sample search and integration, and c.) user-friendly infrastructure to mint PIDs, validate metadata, and provide a sample-specific web landing page (Supplemental Table 1). IGSNs are the only identifier with these characteristics, as they are uniquely governed by an international community organization (IGSN e.V.) with a mission to mint and maintain persistent identifiers for physical samples. The System for Earth Sample Registration (SESAR) is the largest IGSN allocating agent, and enabled us to readily test the process of sample registration and standardizing metadata without first building new infrastructure to mint PIDs, print IGSN barcode labels, and submit and validate metadata. SESAR also provides a persistent sample landing page (e.g. https://www.geosamples.org/profile?igsn=IEBWE000L) with metadata and links to related resources (Lehnert et al. 2019a; Lehnert 2018; Devaraju et al. 2016; Devaraju et al. 2017; McNutt et al. 2016).

Through our comparison of metadata elements in existing sample-related standards and templates (Supplemental Table 3), we concluded that IGSN metadata contains basic information needed, and was therefore sufficient to use in our pilot for standardizing sample metadata.

Sample Identifier and Metadata Testing in the Field

Our pilot test included eight DOE ESS-supported projects that collected field-based samples, including studies of biogeochemical responses to contamination, climate change, or other disturbances (Supplemental Table 4). Project sample types included soil cores, core sections, individual soil samples, sediment, gas, porewater, pond water, river water, leaves, and biofilms. Researchers registered their samples with IGSNs to determine practicalities of using the original SESAR IGSN template (i.e. excel spreadsheet with sample metadata elements for each column and unique sample names/IGSNs for each row)(System for Earth Sample Registration (SESAR), 2020a) in multidisciplinary scientific workflows.

Assigning PIDs and linking related identifiers

A total of 4,485 IGSNs were registered as part of the pilot (Supplemental Table 4). A primary sample for participating projects was often split into multiple subsamples or replicates, and sent to different labs (2-9 labs/user facilities) for numerous analyses (2-23 analyses, Figure 2, Supplemental Table 4; Stegen and Goldman 2018; Chadwick et al. 2020; BERAC 2017; Toyoda et al. 2020). There was universal agreement among researchers that top-level "parent" samples (e.g. soil core), and related "child" samples (e.g. subsections of a soil core) be assigned individual IGSNs. Note that a soil core is a physical parent sample, while in some cases researchers may need to link a set of related samples with no physical parent sample. One example from our test was a set of water samples collected at different depths at a specific point and time in a pond (Figure 3).

Most participants were uncertain whether to assign new IGSNs to subsamples or replicates stored in different containers or split for analyses, particularly when they are essentially considered to be the same sample with the same metadata; many researchers preferred qualifiers/extensions from the same primary IGSN in such cases (Figure 3; Conze et al. 2017). IGSN extensions are currently allowed by request through SESAR IGSN, and are preferred by some users to avoid numerous rounds of IGSN registration and redundant metadata entry. The extensions can allow precise provenance tracking and incorporate additional analytical metadata when subsamples are sent out for a variety of analyses, without requesting new IGSNs. However, this requires users to 1) ensure that their extensions are unique, 2) are restricted to a limited number of additional characters, and 3) that they are batch registered through the IGSN allocating agent with associated metadata, including at least object/sample type, sample name, and the parentIGSN (and ideally all relevant metadata inherited from the parentIGSN). IGSN allocating agents could consider more efficient approaches for registering IGSN subsamples with the same metadata as parentIGSNs, such as adding a metadata field to list subsamples (IGSNs with user-specified extensions), or to have extended IGSNs automatically resolve to the primary IGSN landing page, as done by the ARK identifier system for containment qualifiers (https://wiki.lyrasis.org/display/ARKs/ARK+Identifiers+FAQ).

Researchers also had different opinions on whether related entities (e.g. location) should get an IGSN/PID. In most cases, project-specific, locally unique IDs were sufficient for collection and location IDs. Some researchers assigned IGSNs to wells that were re-sampled over time.

Use of IGSN metadata and template

Much of the IGSN Core Descriptive Metadata is relevant for samples across research domains, but there are key metadata fields and vocabulary terms that are missing or do not accurately describe some ecological samples. We added two essential metadata elements from Minimum Information about any Sequence (MIxS)(i.e. broad environmental context/biome, sample processing; Yilmaz et al, 2011), and added or modified fields based on DarwinCore (i.e. Scientific Name, Depth, and Height fields) to more fully describe ecosystem samples (Figure 5). We concluded that the Environment Ontology (ENVO) includes more relevant terms to describe sample material and environmental context for ecosystem science samples. Because ENVO is used in the MIxS template, it also helps improve interoperability when relating geoscience analyses with 'omics analyses for samples (Table 2), which is often important in ecosystem studies.

IGSN was designed to allow community-specific metadata profiles along with common high-level metadata to support broader interoperability. However, variations across the communities in high-level vocabularies, such as object/sample type and material terms, can inhibit interoperability if the vocabulary terms are not well defined, managed, and linked. We therefore mapped SESAR IGSN terms to ENVO terms for materials. Unlike IGSN vocabulary terms, ENVO terms have specified definitions, PIDs, and are linked to other related terms across many existing ontologies. We also believe that the broader IGSN community could contribute valuable input to the ENVO terms, and benefit from using this ontology or others as they move towards supporting a wider variety of disciplines. We found community agreement that the IGSN Object type terms also need to be revised, and high-level vocabularies will be addressed in the new ESIP Physical Samples Curation Cluster (https://wiki.esipfed.org/Physical_Sample_Curation).

Participants with extensive sampling campaigns found that the spreadsheet format requiring full documentation for each individual sample was impractical. To partially address this, we follow DarwinCore by adding the option of managing metadata using identifiers for higher-level entities (collectionID, locationID, eventID) to help avoid redundant metadata entry. Managing metadata for larger collections of samples by describing sample collections, locations, or events in separate files (see Figure 4) can allow programmatic transfer of relevant metadata to individual samples. However, with regards to applying IGSN metadata to locations we encountered several issues, described in Table 3, as metadata was not intended to fully document site information. We provide additional recommendations in Box 1 that may further improve the efficiency of standardizing sample metadata and/or address practical concerns of researchers.

Sample identifiers for tracking and linking

Researchers generally use their own meaningful sample name for internal sample tracking and individual data analysis workflows (Figure 2); so, both the project-specific sample name and the IGSN should be associated with digital records of the sample. The IGSN, as a globally unique PID, is better suited for automated sample tracking and linking related information over the data life cycle, from field-collection to open-access publication (Guralnick et

al, 2014; Lehnert et al, 2019a). With IGSNs, related samples can be more clearly linked on the sample landing page (e.g. https://www.geosamples.org/profile?igsn=IEWDR000X). Further, specific location or event IDs clarify common relationships for samples and derivatives in a project studying ecological processes at a given location—for example, involving plant litter, leaf, root, soil, and associated 'omics samples.

To most effectively link samples, we recommend that all labs and data systems that generate or store sample data utilize the IGSN or other PID, adding it to metadata templates where relevant. Use of the SESAR API to obtain relevant information about samples can facilitate reuse of metadata across multiple labs or facilities. In theory, the IGSN could be used to automatically add links on the sample landing page to data generated at different facilities; however, no tools are currently available to enable automated linkages.

Improvements are needed to link environmental and associated biological samples. Genomic samples, for example, should be assigned a BioSample number when submitted for sequencing, and linked to the original field-collected sample where relevant (Table 2). There is currently no automated way to link such identifiers, so we recommend providing a full link of the IGSN landing page in the source material ID field in the MIxS template (Table 2).

Discussion

Sample PIDs and metadata in Multidisciplinary Environmental Sciences

We advocate use of IGSNs for ecosystem science samples for a number of reasons. IGSNs are the only PID specifically designed for samples with associated metadata (Lehnert et al, 2019a). IGSN is the only PID backed by an international community of experts, dedicated to identifying, describing, and linking sample data (Lehnert et al, 2011). Participation in the IGSN community will help improve the usefulness of sample PIDs and relevance of associated metadata for multidisciplinary ecosystems science. Additionally, other large national agencies have or plan to adopt IGSNs [e.g. United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), Commonwealth Scientific and Industrial Research Organisation (CSIRO)]. A recently funded effort, iSamples, will improve infrastructure for samples that utilize IGSN and other sample PIDs, and eventually support global search for an even wider variety of sample types (Walls et al. 2020).

Benefits to Data Contributors and Users

Funders of scientific research, such as the US DOE and the National Science Foundation (NSF), require robust data management and publication plans, which should include details for managing and tracking valuable sample data. These data are often not well-described and are missing key information needed to interpret and reuse it, leading to data loss (Michener et al, 1997; Roche et al, 2015; Voytek, 2016). The IGSN-ESS reporting format can assist ecosystem researchers in creating effective sample management plans and preserving their data.

More widespread use of sample PIDs and related metadata will help make sample data more FAIR (Lehnert et al, 2019a; Stall et al, 2019). Standard information to characterize the sample type, location, and date are particularly useful for *finding* relevant data (Poisot et al,

2019; Ponsero et al, 2020). Persistent landing pages for samples allow long-term *access* to sample (meta)data. Use of a controlled vocabulary for key metadata (e.g. sample material and environmental context) helps make data *interoperable* and more easily integrated across datasets. In addition, *reuse* often requires information on collection and processing methods (Poisot et al, 2019). Samples with standard metadata can be more easily shared (i.e. understood and reused) with collaborators, which helps avoid situations where information is lost when people change institutions or retire (Renaut et al, 2018). High-quality published data increasingly helps scientists achieve greater academic recognition, higher citation rates, and can lead to new opportunities for co-authorship and collaboration (Piwowar and Vision, 2013; Whitlock, 2011).

Multidisciplinary ecosystem science often involves complex workflows, and sample PIDs and common metadata provide essential information to help users automatically track samples and add relevant data throughout the sample life cycle. PIDs (such as IGSNs and DOIs) are essential for tracking use of samples and related data over time (Lehnert et al, 2019a; McMurry et al, 2017; Rauber et al, 2016). This provides the foundation to build tools that automatically link and exchange this information across data systems, with no further input from the user after the initial metadata is provided.

Ecosystems research often relies on sample data combined with other data types, such as remote sensing and environmental sensor data, to answer questions about ecosystem response to increasingly rapid global changes (Chadwick et al, 2020; Peters et al, 2014; Serbin et al, 2019; Wu et al, 2020). One limitation is that our standards comparison was focused on sample-related metadata; we need more work towards incorporating standards more suitable for other related entities, such as locations and sensors (Cox, 2017; Esteva et al, 2019; Open Geospatial Consortium Inc., 2010).

More widespread standardization will help reduce the estimated 80% effort currently spent on data wrangling for synthesis work, and enable more efficient data integration and analysis (Renaut et al, 2018). Improved sample data management and reuse will increase the pace of scientific discovery and accelerate new fields of enquiry (Renaut et al, 2018; Roche et al, 2015). Already, publicly available nucleic acid sequences have enabled scientists to build phylogenies and perform comparative genomics studies, and are now essential in community ecology (Webb et al, 2003). Biodiversity records are regularly combined with climate and land use data to predict species distributions, biodiversity, and explore multi-scale ecological patterns (Ball-Damerow et al, 2019; Jetz et al, 2012; Kelling et al, 2009; Renaut et al, 2018). With our multidisciplinary reporting format, we can move beyond infrastructure supporting individual data types, towards efficiently integrating multidisciplinary data to understand ecosystem processes from molecular to global scales.

Conclusions

Summary of IGSN-ESS identifier and metadata recommendations

Many multidisciplinary projects have complicated workflows and need an efficient system for tracking samples as they are sent to different collaborators, labs, user facilities, and published online (Figure 1). Despite growing need and interest, there was previously no straightforward guidance on how to describe sample collections or multidisciplinary samples. We therefore recommend registering samples with IGSNs, using our modified metadata template for ecosystem sciences (IGSN-ESS; Figure 5). The downloadable template, along with complete definitions of all terms, instructions for IGSN registration using IGSN-ESS and providing feedback are detailed in the ESS-DIVE community github repository, and associated data publication (Damerow et al. 2020).

To avoid redundancy in describing samples with the same metadata, we add the optional practice of assigning common sample metadata to a collection, location, or event (Wieczorek et al. 2012; Rocca-Serra et al. 2008; Horsburgh et al. 2016; Diepenbroek et al. 2002). A collection ID provides a flexible way for projects to define common metadata for any set of related samples, while location ID can be used to describe project locations/sites, and event ID can describe metadata for a given sampling event (see Figure 4). These related IDs also provide an unambiguous way to automatically link commonly-related samples. This is particularly important for ecosystem science research, as diverse sample types often need to be clearly linked by specific related identifiers (e.g. location).

For highly-related subsamples with the same metadata, we recommend the option of ID extensions, which could be opaque or meaningful as long as they are unique (Figure 3). It would further improve efficiency of subsample IGSN registration to update the primary IGSN metadata by listing the subsamples or replicates under a “subsample” field, instead of registering them separately. Or the IGSN resolution service could follow the practice of ARKs, where IGSNs with extensions (i.e. containment qualifiers, Supplemental Table 1) automatically resolve to the primary IGSN landing page.

We added or revised fields and vocabulary terms to more accurately describe multidisciplinary samples, and support data linking and reusability (Figure 5). We include controlled vocabularies for relevant subsets of terms from ENVO, (Buttigieg et al. 2016; Damerow et al. 2020), which improves description, search, and integration of a variety of multidisciplinary sample types using key fields (e.g. sample type, sample material, and environmental context; Damerow et al. 2020). We selected terms based on an evaluation of their relevance and likelihood of being used in multiple contexts. We also found that use of ENVO for both local (physiographic feature) and broad (biome) environmental context (e.g. stream ENVO_00000023) is important to fully characterize soil, sediment, and water samples.

Promoting Adoption and other Next Steps

Most ecologists and environmental scientists now understand the importance of data archiving, but struggle to manage data effectively (Renaut et al, 2018; Roche et al, 2015). Removing even trivial barriers can increase the likelihood that researchers will adopt beneficial practices that take effort to achieve (Gardner, 2014). User-friendly guidance and sample metadata templates are an essential step in promoting standard practices that make data publishing, integration, and reuse easier. However, investments are also needed in training programs (Teal et al, 2015), tools to assist with legacy data and analytical instrument systems, and improved data quality management systems that encourage good management practices

throughout the research process (Enke et al, 2012; Freedman et al, 2015). We need tools that translate across existing metadata conventions and use sample and relationship metadata to automatically generate digital resource maps; this could promote adoption by helping users precisely document sample history and linkages to other PIDs and documents (Esteva et al, 2019; Page, 2016). Global sample search (e.g. iSamples Central; Walls et al. 2020), with integrated results, based on key fields (e.g. sample material, location, environmental context, methods, and associated data variables/analyses) would greatly enhance sample data discovery and reuse, and is likely the most effective tool to promote widespread adoption of sample standards (e.g. GBIF; Robertson et al. 2014).

Overcoming complex challenges that require communities to change behavior and provide standardized data will require a coordinated effort, which is best addressed by collaborations of key stakeholders who establish community consensus, enforce guidelines, and help solve problems (Farrell and Simcoe, 2012; Freedman et al, 2015). These stakeholders include a variety of data contributors and users from different scientific domains, as well as laboratory facilities, repositories, funders, and publishers that take part in institutionalizing and rewarding good data management practices (Cousijn et al, 2018; Freedman et al, 2015; Hanson, 2016; Lin and Strasser, 2014). Community coordination on sample reporting conventions and linked cyberinfrastructure will help solve data management problems, expand access pathways, and make our sample data more useful over time.

Data Availability

Data and recommended metadata guidelines generated as part of this work are published in the ESS-DIVE repository, Damerow et al. (2020), and future updates will be managed and available through our community github repository (https://github.com/ess-dive-community/essdive-sample-id-metadata).

Acknowledgements

We greatly appreciate access to SESAR IGSN infrastructure, which allowed us to test use of IGSNs and standardized descriptive metadata with the DOE ESS community. We especially thank Kerstin Lehnert and Sarah Ramdeen for help with sample registration through SESAR, guidance with sample metadata, and for their work with organizing community workshops and meetings regarding samples. JED participated in the IGSN 2040 Steering Committee, and thanks all committee members who contributed to ideas on IGSN value propositions, business planning, and insight regarding direction of IGSN. We also thank Jens Klump for guidance on related work, Chris Mungall for guidance on ontology use and ENVO. We thank other members of the ESS-DIVE team (including Shreyas Cholia, Karen Whitenack) and the National Center for Ecological Analysis and Synthesis (NCEAS) and DataONE (Matt Jones and Chris Jones) who provided information on metadata standards and identifier use across the DataONE network. And, we thank all those who contributed to a variety of community discussions on sample identifiers within the U.S. DOE Biological and Environmental Research community, and broader

groups and discussions across ESIP, RDA, and USGS. Thanks also to Diana Swantek for assistance with final figure designs.

Funding Information

JED, CV, MB, RCO, HE, VH, and DA were funded through the ESS-DIVE repository by the U.S. DOE's Office of Science Biological and Environmental Research under contract number DE-AC02-05CH11231 to LBNL as part of its Earth and Environmental Systems Science Division Data Management program. KSE was supported by the United States Department of Energy contract No. DE-SC0012704 to Brookhaven National Laboratory. RJEA was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research under Contract No. DE-AC02-05CH11231 to LBNL as part of the Terrestrial Ecosystem Science Program. ELB, PS, ZK contributions were supported as part of the Watershed Function Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-AC02-05CH11231. AEG and JCS were supported by the U.S. DOE-BER, as part of BER's Subsurface Biogeochemistry Research (SBR) Program at Pacific Northwest National Laboratory (PNNL), which is operated by Battelle Memorial Institute for the U.S. DOE under Contract No. DE-AC05-76RL01830. ABK, NM, and MZ were supported by the Department of Energy, Office of Science, Biological and Environmental Research, Subsurface Biogeochemical Research program (SCW1053) and performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. RLW's contribution was supported by a National Science Foundation grant 2004562.

Author Contributions

JED, CV, and DA conceived the project. JED conducted the metadata review, led the pilot test, synthesized results and wrote the manuscript. CV and DA supervised the project and provided guidance on its execution. JED, MB, TH, ER, and RW contributed to the comparison of existing sample identifiers, metadata standards and/or reporting conventions. KB, MB, KDC, RJEA, KSE, AEG, VH, ZK, NM, ZP, ER, PS, JCS, PW, MZ participated in field testing and provided feedback on identifiers and metadata. JED and RCO created github documentation. MB, RCO, HE, VH provided input into the reporting format development and documentation as members of the ESS-DIVE repository team. All authors contributed to discussions on the project, reviewed and edited the manuscript.

Competing Interests

The authors declare no competing interests.

References

Avraham, **S**, **Tung**, **C-W**, **Ilic**, **K**, **Jaiswal**, **P**, **Kellogg**, **E A**, **McCouch**, **S**, **Pujar**, **A**, **Reiser**, **L**, **Rhee**, **S Y**, **Sachs**, **M M**, **Schaeffer**, **M**, **Stein**, **L**, **Stevens**, **P**, **Vincent**, **L**, **Zapata**, **F**, and **Ware**, **D** 2008 The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic acids research*, 36(Database issue): D449–54. DOI: https://doi.org/10.1093/nar/gkm908

Ball-Damerow, **J E**, **Brenskelle**, **L**, **Barve**, **N**, **Soltis**, **P S**, **Sierwald**, **P**, **Bieler**, **R**, **LaFrance**, **R**, **Ariño**, **A H**, and **Guralnick**, **R P** 2019 Research applications of primary biodiversity databases in the digital age. *PloS one*, 14(9): e0215794. DOI: https://doi.org/10.1371/journal.pone.0215794

Beck, **M W**, **O'Hara**, **C**, **Stewart Lowndes**, **J S**, **D Mazor**, **R**, **Theroux**, **S**, **J Gillett**, **D**, **Lane**, **B**, and **Gearheart**, **G** 2020 The importance of open science for biological assessment of aquatic environments. *PeerJ*, 8: e9539. DOI: https://doi.org/10.7717/peerj.9539

Biological and Environmental Research Advisory Committee (BERAC) 2017 Grand Challenges for Biological and Environmental Research: Progress and Future Vision 2017: A Report from the Biological and Environmental Research Advisory Committee (No. DOE/SC–0190). BERAC Subcommittee on Grand Research Challenges for Biological and Environmental Research. Available at https://genomicscience.energy.gov/BERfiles/BERAC-2017-Grand-Challenges-Report.pdf

Buttigieg, **P L**, **Pafilis**, **E**, **Lewis**, **S E**, **Schildhauer**, **M P**, **Walls**, **R L**, and **Mungall**, **C J** 2016 The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of biomedical semantics*, 7(1): 57. DOI: https://doi.org/10.1186/s13326-016-0097-6

Chadwick, **K D**, **Brodrick**, **P G**, **Grant**, **K**, **Goulden**, **T**, **Henderson**, **A**, **Falco**, **N**, **Wainwright**, **H**, **Williams**, **K H**, **Bill**, **M**, **Breckheimer**, **I**, **Brodie**, **E L**, **Steltzer**, **H**, **Williams**, **C F R**, **Blonder**, **B**, **Chen**, **J**, **Dafflon**, **B**, **Damerow**, **J**, **Hancher**, **M**, **Khurram**, **A**, **Lamb**, **J**, **Lawrence**, **C R**, **McCormick**, **M**, **Musinsky**, **J**, **Pierce**, **S**, **Polussa**, **A**, **Hastings Porro**, **M**, **Scott**, **A**, **Singh**, **H W**, **Sorensen**, **P O**, **Varadharajan**, **C**, **Whitney**, **B**, and **Maher**, **K** 2020 Integrating airborne remote sensing and field campaigns for ecology and Earth system science. *Methods in ecology and evolution / British Ecological Society*, 1: 83. DOI: https://doi.org/10.1111/2041-210X.13463

Chase, **J H**, **Bolyen**, **E**, **Rideout**, **J R**, and **Caporaso**, **J G** 2016 cual-id: Globally Unique, Correctable, and Human-Friendly Sample Identifiers for Comparative Omics Studies. *mSystems*, 1(1): e00010–15. DOI: https://doi.org/10.1128/mSystems.00010-15

Conze, **R**, **Lorenz**, **H**, **Ulbricht**, **D**, **Elger**, **K**, and **Gorgas**, **T** 2017 Utilizing the International Geo Sample Number Concept in Continental Scientific Drilling During ICDP Expedition COSC-1. *Data Science Journal*, 16(0): 2. DOI: https://doi.org/10.5334/dsj-2017-002

Cordeiro, **A L**, **Norby**, **R J**, **Andersen**, **K M**, **Valverde-Barrantes**, **O**, **Fuchslueger**, **L**, **Oblitas**, **E**, **Hartley**, **I P**, **Iversen**, **C M**, **Gonçalves**, **N B**, **Takeshi**, **B**, **Lapola**, **D M**, and **Quesada**, **C A** 2020 Fine-root dynamics vary with soil depth and precipitation in a low-nutrient tropical forest in

the Central Amazonia. *Plant-Environment Interactions*, 1(1): 3–16. DOI: https://doi.org/10.1002/pei3.10010

Cousijn, **H**, **Kenall**, **A**, **Ganley**, **E**, **Harrison**, **M**, **Kernohan**, **D**, **Lemberger**, **T**, **Murphy**, **F**, **Polischuk**, **P**, **Taylor**, **S**, **Martone**, **M**, and **Clark**, **T** 2018 A data citation roadmap for scientific publishers. *Scientific Data*, 5(1): 1–11. DOI: https://doi.org/10.1038/sdata.2018.259

Cox, **S J D** 2017 Ontology for observations and sampling features, with alignments to existing models. *Semantic Web*, 8(3): 453–470. DOI: https://doi.org/10.3233/SW-160214

Damerow, **J**, **Varadharajan**, **C**, **Boye**, **K**, **Brodie**, **E**, **Chadwick**, **D**, **Cholia**, **S**, **Elbashandy**, **H**, **Ely**, **K**, **Goldman**, **A**, **Hendrix**, **V**, **Jones**, **C**, **Jones**, **M**, **Kakalia**, **Z**, **Kemner**, **K**, **Kersting**, **A**, **Maher**, **K**, **Merino**, **N**, **O'Brien**, **F**, **Perzan**, **Z**, **Robles**, **E**, **Snavely**, **C**, **Sorensen**, **P**, **Stegen**, **J**, **Weisenhorn**, **P**, **Whitenack**, **K**, **Zavarin**, **M**, and **Agarwal**, **D** 2020 Sample Identifiers and Metadata Reporting Format for Environmental Systems Science. *Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE)*. DOI: https://doi.org/10.15485/1660470

Darwin Core Task Group 2014 Darwin Core: 2014-11-08. Biodiversity Information Standards (TDWG). DOI: https://doi.org/10.5281/zenodo.12694

DataCite Metadata Working Group 2019 DataCite Metadata Schema for the Publication and Citation of Research Data. DataCite e.V. Available at http://doi.org/10.5438/rv0g-av03 [Last accessed 16 September 2020].

DCMI Usage Board 2020 Dublin Core Metadata Initiative (DCMI) Metadata Terms. Available at https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Devaraju, **A**, **Klump**, **J**, **Cox**, **S J D**, and **Golodoniuc**, **P** 2016 Representing and publishing physical sample descriptions. *Computers & geosciences*, 96: 1–10. DOI: https://doi.org/10.1016/j.cageo.2016.07.018

Devaraju, **A**, **Klump**, **J**, **Tey**, **V**, **Fraser**, **R**, **Cox**, **S**, and **Wyborn**, **L** 2017 A Digital Repository for Physical Samples: Concepts, Solutions and Management. In: Kamps, J, Tsakonas, G, Manolopoulos, Y, Iliadis, L, and Karydis, I (eds.), *Research and Advanced Technology for Digital Libraries*. Springer International Publishing. pp. 74–85.

Diepenbroek, M, Grobe, H, Reinke, M, Schindler, U, Schlitzer, R, Sieger, R, & Wefer, G 2002 PANGAEA - an information system for environmental sciences. Computers & Geosciences, 28(10), 1201–1210. DOI: https://doi.org/10.1016/S0098-3004(02)00039-0

Duerr, **R E**, **Downs**, **R R**, **Tilmes**, **C**, **Barkstrom**, **B**, **Lenhardt**, **W C**, **Glassy**, **J**, **Bermudez**, **L E**, and **Slaughter**, **P** 2011 On the utility of identification schemes for digital earth science data: an assessment and recommendations. *Earth Science Informatics*, 4(3): 139. DOI: https://doi.org/10.1007/s12145-011-0083-6

Enke, **N**, **Thessen**, **A**, **Bach**, **K**, **Bendix**, **J**, **Seeger**, **B**, and **Gemeinholzer**, **B** 2012 The user's view on biodiversity data sharing — Investigating facts of acceptance and requirements to realize a sustainable use of research data —. *Ecological informatics*, 11: 25–33. DOI: https://doi.org/10.1016/j.ecoinf.2012.03.004

Esteva, **M**, **Walls**, **R L**, **Magill**, **A B**, **Xu**, **W**, **Huang**, **R**, **Carson**, **J**, and **Song**, **J** 2019 Identifier Services: Modeling and Implementing Distributed Data Management in Cyberinfrastructure. *Data and Information Management*, 3(1): 26–39. DOI: https://doi.org/10.2478/dim-2019-0002

Farrell, **J** and **Simcoe**, **T** 2012 Four Paths to Compatibility. In: Martin Peitz And (ed.), *The Oxford Handbook of the Digital Economy*. Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780195397840.013.0002

Ferguson, **C**, **McEntrye**, **J**, **Bunakov**, **V**, **Lambert**, **S**, **van der Sandt**, **S**, **Kotarski**, **R**, **Stewart**, **S**, **MacEwan**, **A**, **Fenner**, **M**, **Cruse**, **P**, **van Horik**, **R**, **Dohna**, **T**, **Koop-Jacobsen**, **K**, **Schindler**, **U**, and **McCafferty**, **S** 2018 Survey of Current PID Services Landscape. FREYA Consortium.

Field, **D**, **Amaral-Zettler**, **L**, **Cochrane**, **G**, **Cole**, **J R**, **Dawyndt**, **P**, **Garrity**, **G M**, **Gilbert**, **J**, **Glöckner**, **F O**, **Hirschman**, **L**, **Karsch-Mizrachi**, **I**, **Klenk**, **H-P**, **Knight**, **R**, **Kottmann**, **R**, **Kyrpides**, **N**, **Meyer**, **F**, **San Gil**, **I**, **Sansone**, **S-A**, **Schriml**, **L M**, **Sterk**, **P**, **Tatusova**, **T**, **Ussery**, **D W**, **White**, **O**, and **Wooley**, **J** 2011 The Genomic Standards Consortium. *PLoS biology*, 9(6): e1001088. DOI: https://doi.org/10.1371/journal.pbio.1001088

Freedman, **L P**, **Cockburn**, **I M**, and **Simcoe**, **T S** 2015 The Economics of Reproducibility in Preclinical Research. *PLoS biology*, 13(6): e1002165. DOI: https://doi.org/10.1371/journal.pbio.1002165

Gaiji, **S**, **Chavan**, **V**, **Ariño**, **A H**, **Otegui**, **J**, **Hobern**, **D**, **Sood**, **R**, and **Robles**, **E** 2013 Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. *Biodiversity Informatics*, 8(2). DOI: https://doi.org/10.17161/bi.v8i2.4124

Gardner, **T** 2014 A swan in the making. *Science*, 345(6199): 855–855. DOI: https://doi.org/10.1126/science.1259740

Goldstein, **S**, **Lehnert**, **K**, and **Hofmann**, **A** 2014 Requirements for the Publication of Geochemical Data. Available at http://doi.iedadata.org/100426 [Last accessed 11 February 2019].

Guralnick, **R**, **Conlin**, **T**, **Deck**, **J**, **Stucky**, **B J**, and **Cellinese**, **N** 2014 The Trouble with Triplets in Biodiversity Informatics: A Data-Driven Case against Current Identifier Practices. *PloS one*, 9(12): e114069. DOI: https://doi.org/10.1371/journal.pone.0114069

Guralnick, **R P**, **Cellinese**, **N**, **Deck**, **J**, **Pyle**, **R L**, **Kunze**, **J**, **Penev**, **L**, **Walls**, **R**, **Hagedorn**, **G**, **Agosti**, **D**, **Wieczorek**, **J**, **Catapano**, **T**, and **Page**, **R** 2015 Community Next Steps for Making Globally Unique Identifiers Work for Biocollections Data. *ZooKeys*, 494: 133–154. DOI: https://doi.org/10.3897/zookeys.494.9352

Hanson, **B** 2016 AGU Opens Its Journals to Author Identifiers. Available at https://eos.org/agu-news/agu-opens-its-journals-to-author-identifiers [Last accessed 28 September 2020].

Horsburgh, **J S**, **Aufdenkampe**, **A K**, **Mayorga**, **E**, **Lehnert**, **K A**, **Hsu**, **L**, **Song**, **L**, **Jones**, **A S**, **Damiano**, **S G**, **Tarboton**, **D G**, **Valentine**, **D**, **Zaslavsky**, **I**, and **Whitenack**, **T** 2016 Observations Data Model 2: A community information model for spatially discrete Earth observations. *Environmental Modelling & Software*, 79: 55–74. DOI:

https://doi.org/10.1016/j.envsoft.2016.01.010

Jetz, **W**, **Thomas**, **G H**, **Joy**, **J B**, **Hartmann**, **K**, and **Mooers**, **A O** 2012 The global diversity of birds in space and time. *Nature*, 491(7424): 444–448. DOI: https://doi.org/10.1038/nature11631

Kelling, **S**, **Hochachka**, **W M**, **Fink**, **D**, **Riedewald**, **M**, **Caruana**, **R**, **Ballard**, **G**, and **Hooker**, **G** 2009 Data-intensive Science: A New Paradigm for Biodiversity Studies. *Bioscience*, 59(7): 613–620. DOI: https://doi.org/10.1525/bio.2009.59.7.12

Klump, **J**, **Huber**, **R**, and **Diepenbroek**, **M** 2016 DOI for geoscience data - how early practices shape present perceptions. *Earth Science Informatics*, 9(1): 123–136. DOI: https://doi.org/10.1007/s12145-015-0231-5

Lehnert, **K** 2018 *IGSN: Toward a Mature and Generic Persistent Identifier for Samples*. Available at https://www.slideshare.net/klehnert/igsn-toward-a-mature-and-generic-persistent-identifier-for-samples [Last accessed 25 January 2019].

Lehnert, **K A**, **Klump**, **J**, **Arko**, **R A**, **Bristol**, **S**, **Buczkowski**, **B**, **Chan**, **C**, **Chan**, **S**, **Conze**, **R**, **Cox**, **S J**, **Habermann**, **T**, **Hangsterfer**, **A**, **Hsu**, **L**, **Milan**, **A**, **Miller**, **S P**, **Noren**, **A J**, **Richard**, **S M**, **Valentine**, **D W**, **Whitenack**, **T**, **Wyborn**, **L A**, and **Zaslavsky**, **I** 2011 IGSN e.V.: Registration and Identification Services for Physical Samples in the Digital Universe. *AGU Fall Meeting Abstracts*, 13: IN13B–1324. Available at http://adsabs.harvard.edu/abs/2011AGUFMIN13B1324L [Last accessed 1 March 2019].

Lehnert, **K**, **Klump**, **J**, **Wyborn**, **L**, and **Ramdeen**, **S** 2019a Persistent, Global, Unique: The three key requirements for a trusted identifier system for physical samples. *Biodiversity Information Science and Standards*, 3: e37334. DOI: https://doi.org/10.3897/biss.3.37334

Lehnert, **K**, **Wyborn**, **L**, and **Klump**, **J** 2019b FAIR Geoscientific Samples and Data Need International Collaboration. *Acta Geologica Sinica - English Edition*, 93(S3): 32–33. DOI: https://doi.org/10.1111/1755-6724.14236

Lin, **J** and **Strasser**, **C** 2014 Recommendations for the Role of Publishers in Access to Data. *PLoS biology*, 12(10): e1001975. DOI: https://doi.org/10.1371/journal.pbio.1001975

Malik, **A A**, **Martiny**, **J B H**, **Brodie**, **E L**, **Martiny**, **A C**, **Treseder**, **K K**, and **Allison**, **S D** 2020 Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *The ISME journal*, 14(1): 1–9. DOI: https://doi.org/10.1038/s41396-019-0510-0

McMurry, **J A**, **Juty**, **N**, **Blomberg**, **N**, **Burdett**, **T**, **Conlin**, **T**, **Conte**, **N**, **Courtot**, **M**, **Deck**, **J**, **Dumontier**, **M**, **Fellows**, **D K**, **Gonzalez-Beltran**, **A**, **Gormanns**, **P**, **Grethe**, **J**, **Hastings**, **J**, **Hériché**, **J-K**, **Hermjakob**, **H**, **Ison**, **J C**, **Jimenez**, **R C**, **Jupp**, **S**, **Kunze**, **J**, **Laibe**, **C**, **Le Novère**, **N**, **Malone**, **J**, **Martin**, **M J**, **McEntyre**, **J R**, **Morris**, **C**, **Muilu**, **J**, **Müller**, **W**, **Rocca-Serra**, **P**, **Sansone**, **S-A**, **Sariyar**, **M**, **Snoep**, **J L**, **Soiland-Reyes**, **S**, **Stanford**, **N J**, **Swainston**, **N**, **Washington**, **N**, **Williams**, **A R**, **Wimalaratne**, **S M**, **Winfree**, **L M**, **Wolstencroft**, **K**, **Goble**, **C**, **Mungall**, **C J**, **Haendel**, **M A**, and **Parkinson**, **H** 2017 Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS biology*, 15(6). DOI: https://doi.org/10.1371/journal.pbio.2001414

McNutt, **M**, **Lehnert**, **K**, **Hanson**, **B**, **Nosek**, **B A**, **Ellison**, **A M**, and **King**, **J L** 2016 Liberating

field science samples and data. *Science*, 351(6277): 1024–1026. DOI: https://doi.org/10.1126/science.aad7048

Michener, **W K** 2015 Ecological data sharing. *Ecological informatics*, 29: 33–44. DOI: https://doi.org/10.1016/j.ecoinf.2015.06.010

Michener, **W K**, **Brunt**, **J W**, **Helly**, **J J**, **Kirchner**, **T B**, and **Stafford**, **S G** 1997 Nongeospatial Metadata for the Ecological Sciences. *Ecological applications: a publication of the Ecological Society of America*, 7(1): 330–342. DOI: https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2

Open Geospatial Consortium Inc. 2010 ISO 19156:2011 - Geographic information -- Observations and measurements. International Organization for Standardization. DOI: https://doi.org/10.13140/2.1.1142.3042

Page, **R** 2016 Towards a biodiversity knowledge graph. *Research Ideas and Outcomes*, 2: e8767. DOI: https://doi.org/10.3897/rio.2.e8767

Peters, **D P C**, **Loescher**, **H W**, **SanClements**, **M D**, and **Havstad**, **K M** 2014 Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology. *Ecosphere* , 5(3): art29. DOI: https://doi.org/10.1890/ES13-00295.1

Rocca-Serra, P, Sansone, S-A, Brandizi, M et al. 2008 ISA-TAB Specification Documentation. Available at http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf

Piwowar, **H A** and **Vision**, **T J** 2013 Data reuse and the open data citation advantage. *PeerJ*, 1: e175. DOI: https://doi.org/10.7717/peerj.175

Poisot, **T**, **Bruneau**, **A**, **Gonzalez**, **A**, **Gravel**, **D**, and **Peres-Neto**, **P** 2019 Ecological Data Should Not Be So Hard to Find and Reuse. *Trends in ecology & evolution*, 34(6): 494–496. DOI: https://doi.org/10.1016/j.tree.2019.04.005

Ponsero, **A J**, **Bomhoff**, **M**, **Blumberg**, **K**, **Youens-Clark**, **K**, **Herz**, **N M**, **Wood-Charlson**, **E M**, **Delong**, **E F**, and **Hurwitz**, **B L** 2020 Planet Microbe: a platform for marine microbiology to discover and analyze interconnected ‘omics and environmental data. *Nucleic acids research*. DOI: https://doi.org/10.1093/nar/gkaa637

Rauber, **A**, **Asmi**, **A**, **Van Uytvanck**, **D**, and **Proell**, **S** 2016 Data Citation of Evolving Data: Recommendations of the RDA Working Group on Data Citation (WGDC). *Research Data Alliance*. DOI: https://doi.org/10.15497/rda00016

Reddy, **T B K**, **Thomas**, **A D**, **Stamatis**, **D**, **Bertsch**, **J**, **Isbandi**, **M**, **Jansson**, **J**, **Mallajosyula**, **J**, **Pagani**, **I**, **Lobos**, **E A**, and **Kyrpides**, **N C** 2015 The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic acids research*, 43(Database issue): D1099–106. DOI: https://doi.org/10.1093/nar/gku950

Renaut, **S**, **Budden**, **A E**, **Gravel**, **D**, **Poisot**, **T**, and **Peres-Neto**, **P** 2018 Management, Archiving, and Sharing for Biologists and the Role of Research Institutions in the Technology-Oriented Age. *Bioscience*, 68(6): 400–411. DOI: https://doi.org/10.1093/biosci/biy038

Robertson, **T**, **Döring**, **M**, **Guralnick**, **R**, **Bloom**, **D**, **Wieczorek**, **J**, **Braak**, **K**, **Otegui**, **J**, **Russell**, **L**, and **Desmet**, **P** 2014 The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. *PloS one*, 9(8): e102623. DOI: https://doi.org/10.1371/journal.pone.0102623

Roche, **D G**, **Kruuk**, **L E B**, **Lanfear**, **R**, and **Binning**, **S A** 2015 Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLoS biology*, 13(11): e1002295. DOI: https://doi.org/10.1371/journal.pbio.1002295

Serbin, **S P**, **Wu**, **J**, **Ely**, **K S**, **Kruger**, **E L**, **Townsend**, **P A**, **Meng**, **R**, **Wolfe**, **B T**, **Chlus**, **A**, **Wang**, **Z**, and **Rogers**, **A** 2019 From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. *The New phytologist*, 224(4): 1557–1568. DOI: https://doi.org/10.1111/nph.16123

Soranno, **P A** and **Schimel**, **D S** 2014 Macrosystems ecology: big data, big ecology. *Frontiers in ecology and the environment*, 12(1): 3–3. DOI: https://doi.org/10.1890/1540-9295-12.1.3

Stall, **S**, **Yarmey**, **L**, **Cutcher-Gershenfeld**, **J**, **Hanson**, **B**, **Lehnert**, **K**, **Nosek**, **B**, **Parsons**, **M**, **Robinson**, **E**, and **Wyborn**, **L** 2019 Make scientific data FAIR. *Nature*, 570(7759): 27. DOI: https://doi.org/10.1038/d41586-019-01720-7

Stegen, **J C** and **Goldman**, **A E** 2018 WHONDRS: a Community Resource for Studying Dynamic River Corridors. *mSystems*, 3(5): e00151–18. DOI: https://doi.org/10.1128/mSystems.00151-18

System for Earth Sample Registration (SESAR) 2020a SESAR Batch Registration Quick Guide. DOI: https://doi.org/10.5281/zenodo.3874923

System for Earth Sample Registration (SESAR) 2020b SESAR XML Schema for samples. DOI: https://doi.org/10.5281/zenodo.3875531

Teal, **T K**, **Cranston**, **K A**, **Lapp**, **H**, **White**, **E**, **Wilson**, **G**, **Ram**, **K**, and **Pawlik**, **A** 2015 Data Carpentry: Workshops to Increase Data Literacy for Researchers, 10(1): 135–143. DOI: https://doi.org/10.2218/ijdc.v10i1.351

Toyoda, **J G**, **Goldman**, **A E**, **Chu**, **R K**, **Danczak**, **R E**, and **Daly**, **R A** 2020 WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Surface Water FTICR-MS, NPOC, and Stable Isotopes. Available at https://data.ess-dive.lbl.gov/view/doi:10.15485/1603775 [Last accessed 16 November 2020].

Treseder, **K K**, **Balser**, **T C**, **Bradford**, **M A**, **Brodie**, **E L**, **Dubinsky**, **E A**, **Eviner**, **V T**, **Hofmockel**, **K S**, **Lennon**, **J T**, **Levine**, **U Y**, **MacGregor**, **B J**, **Pett-Ridge**, **J**, and **Waldrop**, **M P** 2012 Integrating microbial ecology into ecosystem models: challenges and priorities. *Biogeochemistry*, 109(1): 7–18. DOI: https://doi.org/10.1007/s10533-011-9636-5

Treloar, A, Klump, J 2019 Updating the Data Curation Continuum: not just Data, still focused on Curation, more about Domains. International Journal of Digital Curation, 14(1): 87–101. DOI: https://doi.org/10.2218/ijdc.v14i1.643

Varadharajan, **C**, **Cholia**, **S**, **Snavely**, **C**, **Hendrix**, **V**, **Procopiou**, **C**, **Riley**, **W**, and **Agarwal**, **D A** 2018 Launching an Accessible Archive of Environmental Data. *Eos* , 100. DOI:

https://doi.org/10.1029/2019EO111263.

Voytek, **B** 2016 The Virtuous Cycle of a Data Ecosystem. *PLoS computational biology*, 12(8): e1005037. DOI: https://doi.org/10.1371/journal.pcbi.1005037

Walls, **R L**, **Deck**, **J**, **Guralnick**, **R**, **Baskauf**, **S**, **Beaman**, **R**, **Blum**, **S**, **Bowers**, **S**, **Buttigieg**, **P L**, **Davies**, **N**, **Endresen**, **D**, **Gandolfo**, **M A**, **Hanner**, **R**, **Janning**, **A**, **Krishtalka**, **L**, **Matsunaga**, **A**, **Midford**, **P**, **Morrison**, **N**, **Tuama**, **É Ó**, **Schildhauer**, **M**, **Smith**, **B**, **Stucky**, **B J**, **Thomer**, **A**, **Wieczorek**, **J**, **Whitacre**, **J**, and **Wooley**, **J** 2014 Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. *PloS one*, 9(3): e89606. DOI: https://doi.org/10.1371/journal.pone.0089606

Walls, R, Davies, N, Kansa, S; Kunze, J, Lehnert, K, Vieglais, D 2020 Building transdisciplinary infrastructure for natural history material samples with the Internet of Samples (iSamples). Zenodo. DOI: https://zenodo.org/record/4002440

Weart, **S** 2013 Rise of interdisciplinary research on climate. *Proceedings of the National Academy of Sciences of the United States of America*, 110 Suppl 1: 3657–3664. DOI: https://doi.org/10.1073/pnas.1107482109

Webb, **C O**, **Ackerly**, **D D**, **McPeek**, **M A**, and **Donoghue**, **M J** 2003 Phylogenies and Community Ecology. DOI: https://doi.org/10.1146/annurev.ecolsys.33.010802.150448

Whitlock, **M C** 2011 Data archiving in ecology and evolution: best practices. *Trends in ecology & evolution*, 26(2): 61–65. DOI: https://doi.org/10.1016/j.tree.2010.11.006

Wieczorek, **J**, **Bloom**, **D**, **Guralnick**, **R**, **Blum**, **S**, **Döring**, **M**, **Giovanni**, **R**, **Robertson**, **T**, and **Vieglais**, **D** 2012 Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PloS one*, 7(1): e29715. DOI: https://doi.org/10.1371/journal.pone.0029715

Wilkinson, **M D**, **Dumontier**, **M**, **Aalbersberg**, **I J**, **Appleton**, **G**, **Axton**, **M**, **Baak**, **A**, **Blomberg**, **N**, **Boiten**, **J-W**, **da Silva Santos**, **L B**, **Bourne**, **P E**, **Bouwman**, **J**, **Brookes**, **A J**, **Clark**, **T**, **Crosas**, **M**, **Dillo**, **I**, **Dumon**, **O**, **Edmunds**, **S**, **Evelo**, **C T**, **Finkers**, **R**, **Gonzalez-Beltran**, **A**, **Gray**, **A J G**, **Groth**, **P**, **Goble**, **C**, **Grethe**, **J S**, **Heringa**, **J**, **'t Hoen**, **P A C**, **Hooft**, **R**, **Kuhn**, **T**, **Kok**, **R**, **Kok**, **J**, **Lusher**, **S J**, **Martone**, **M E**, **Mons**, **A**, **Packer**, **A L**, **Persson**, **B**, **Rocca-Serra**, **P**, **Roos**, **M**, **van Schaik**, **R**, **Sansone**, **S-A**, **Schultes**, **E**, **Sengstag**, **T**, **Slater**, **T**, **Strawn**, **G**, **Swertz**, **M A**, **Thompson**, **M**, **van der Lei**, **J**, **van Mulligen**, **E**, **Velterop**, **J**, **Waagmeester**, **A**, **Wittenburg**, **P**, **Wolstencroft**, **K**, **Zhao**, **J**, and **Mons**, **B** 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. DOI: https://doi.org/10.1038/sdata.2016.18

Wu, **J**, **Rogers**, **A**, **Albert**, **L P**, **Ely**, **K**, **Prohaska**, **N**, **Wolfe**, **B T**, **Oliveira**, **R C**, Jr, **Saleska**, **S R**, and **Serbin**, **S P** 2019 Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests. *The New phytologist*, 224(2): 663–674. DOI: https://doi.org/10.1111/nph.16029

Wu, **J**, **Serbin**, **S P**, **Ely**, **K S**, **Wolfe**, **B T**, **Dickman**, **L T**, **Grossiord**, **C**, **Michaletz**, **S T**, **Collins**, **A D**, **Detto**, **M**, **McDowell**, **N G**, **Wright**, **S J**, and **Rogers**, **A** 2020 The response of stomatal conductance to seasonal drought in tropical forests. *Global change biology*, 26(2): 823–839. DOI: https://doi.org/10.1111/gcb.14820

Yilmaz, **P**, **Kottmann**, **R**, **Field**, **D**, **Knight**, **R**, **Cole**, **J R**, **Amaral-Zettler**, **L**, **Gilbert**, **J A**, **Karsch-Mizrachi**, **I**, **Johnston**, **A**, **Cochrane**, **G**, **Vaughan**, **R**, **Hunter**, **C**, **Park**, **J**, **Morrison**, **N**, **Rocca-Serra**, **P**, **Sterk**, **P**, **Arumugam**, **M**, **Bailey**, **M**, **Baumgartner**, **L**, **Birren**, **B W**, **Blaser**, **M J**, **Bonazzi**, **V**, **Booth**, **T**, **Bork**, **P**, **Bushman**, **F D**, **Buttigieg**, **P L**, **Chain**, **P S G**, **Charlson**, **E**, **Costello**, **E K**, **Huot-Creasy**, **H**, **Dawyndt**, **P**, **DeSantis**, **T**, **Fierer**, **N**, **Fuhrman**, **J A**, **Gallery**, **R E**, **Gevers**, **D**, **Gibbs**, **R A**, **Gil**, **I S**, **Gonzalez**, **A**, **Gordon**, **J I**, **Guralnick**, **R**, **Hankeln**, **W**, **Highlander**, **S**, **Hugenholtz**, **P**, **Jansson**, **J**, **Kau**, **A L**, **Kelley**, **S T**, **Kennedy**, **J**, **Knights**, **D**, **Koren**, **O**, **Kuczynski**, **J**, **Kyrpides**, **N**, **Larsen**, **R**, **Lauber**, **C L**, **Legg**, **T**, **Ley**, **R E**, **Lozupone**, **C A**, **Ludwig**, **W**, **Lyons**, **D**, **Maguire**, **E**, **Methé**, **B A**, **Meyer**, **F**, **Muegge**, **B**, **Nakielny**, **S**, **Nelson**, **K E**, **Nemergut**, **D**, **Neufeld**, **J D**, **Newbold**, **L K**, **Oliver**, **A E**, **Pace**, **N R**, **Palanisamy**, **G**, **Peplies**, **J**, **Petrosino**, **J**, **Proctor**, **L**, **Pruesse**, **E**, **Quast**, **C**, **Raes**, **J**, **Ratnasingham**, **S**, **Ravel**, **J**, **Relman**, **D A**, **Assunta-Sansone**, **S**, **Schloss**, **P D**, **Schriml**, **L**, **Sinha**, **R**, **Smith**, **M I**, **Sodergren**, **E**, **Spor**, **A**, **Stombaugh**, **J**, **Tiedje**, **J M**, **Ward**, **D V**, **Weinstock**, **G M**, **Wendel**, **D**, **White**, **O**, **Whiteley**, **A**, **Wilke**, **A**, **Wortman**, **J R**, **Yatsunenko**, **T**, and **Glöckner**, **F O** 2011 Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature biotechnology*, 29(5): 415–420. DOI: https://doi.org/10.1038/nbt.1823

Figures

Figure 1. Tracking interdisciplinary samples throughout the cycle of field collection, transport to collaborators and other labs, various analyses, and digital records.

Figure 2. Sample journey map, using the sample PID and metadata to document sample history and link related samples in the WHONDRS project (Stegen and Goldman 2018; Toyoda et al. 2020).

Figure 3. Options for assigning IDs to sets or chains of highly related samples and subsamples. There is uncertainty among domain scientists about whether to assign new PIDs to subsamples. Based on our pilot test feedback, options 2 and 3 are most efficient for soil cores and water samples, respectively. Relationship metadata can be inferred from the type of ID (e.g. collection or site ID) and the order of Parent IGSNs, and assists machine reconstruction of the sampling hierarchy from original feature or sample through subsequent child samples.

Figure 4: Example of using **related** identifiers to link related samples and information. **Related** identifiers are listed in blue. All metadata can be provided at the sample level or by providing separate files (depicted as boxes) for higher-level collections of samples, sampling events, methods, and/or locations. When providing separate spreadsheet files, each file (e.g. locations file) contains a row for each unique related identifier (e.g. location ID), with the associated metadata fields (e.g. location description) as columns. Unique identifiers for these related, higher-level entities then allow associating relevant metadata (e.g. latitude and longitude) with individual samples. This practice is flexible and optional, depending on data management needs and preferences.

Figure 5: Sample metadata for Environmental Systems Sciences (IGSN-ESS). Each sample metadata element is listed under a general category of information. Required fields are marked with an asterisk*. Fields added to IGSN metadata or revised from Darwin Core (DwC), MIxS, Environment Ontology (ENVO), Biological Collections Ontology (BCO), Plant Ontology (PO) are indicated in parentheses.

Tables

Identifier Type	Identifier Example	Scope
ARK	ark:/12148/btv1b8449691v	Flexible
URN	urn:catalog:UMMZ:Mammals:171041	Flexible
HTTP URI	http://data.rbge.org.uk/herb/E00115694	Flexible
DOI	10.7299/X7VQ32SJ	Flexible, mostly papers and datasets
UUID	EF0A4D3E-702F-4882-81B8-CA737AEB7B28	Flexible
IGSN	IGSN: IECUR0002	Geoscience, working to become general physical sample identifier
CETAF URI, based on HTTP URI	http://data.rbge.org.uk/herb/E00421503	Species Occurrence, Specimens from CETAF institutions
RRID	RRID:MGI:5630441	Biomedical Research Resources
BioSample accession number	SAMN03983893	Biological source materials used in experimental assays

Table 1. Examples of PIDs that have been used for samples, modified from Guralnick et al. (2015). Acronyms: ARK=Archival Resource Keys, URN=Uniform Resource Name, URI=Uniform Resource Identifier, DOI=Digital Object Identifier, UUID=Universally Unique Identifier, IGSN=International GeoSample Number, CETAF=Consortium of the European Taxonomic Facilities, RRID=Research Resource Identifier

IGSN field	**MIxS/MIMs field**
IGSN	Source material ID (can include the full link to sample landing page)
Material	Environmental medium* = ENVO
Related to Material	organism (e.g. soil metagenome)
Physiographic feature	local scale environmental context* = ENVO
N/A	broad scale environmental context* = ENVO
Country	geographic location (country or region) = GAZ
N/A	sample material processing

Table 2. Mapping of key fields to promote interoperability between geoscience (IGSN) and associated metagenomic samples (BioSample). Minimum Information about Any Sequence (MIxS) / Minimum Information about any Metagenomic Sequence (MIMS) templates require or encourage use of the Environment Ontology (ENVO) to describe environmental context and materials, and the GAZETTEER ontology (GAZ) for place names.

Location ID	If there is a project-specific site/location name, you must currently provide this in the free-text location description field. We therefore added LocationID as a field, which can be associated with metadata and does not need to be globally unique. Sample metadata contains location fields, but is not intended to fully describe sites/location information.
Location Hierarchies	We do not address a standard way to represent complex location hierarchies (e.g. basins, watersheds, wells, depths within wells), which is needed but is out of scope for the current effort.
Plot Name	Many projects are located in remote areas where GPS coordinates are not reliable and yet specific locations are necessary. Therefore, plots are formally defined and distance from specific points documented in the field using a relative reference system. Currently, users must describe this within the Location Description metadata field.
Uncertainty or precision of geographic coordinates	We could add a metadata field to provide detail on the uncertainty in the geographic coordinates, as done in DarwinCore. However, we found that participants sometimes do not have this information. Certain instruments (i.e. smart phones) do not provide an easy way to specify uncertainty. It may therefore be more efficient to simply indicate the specific instrument used to provide information on the likely uncertainty or precision of the coordinates. Additional terms are needed to specify instrument used.
Sampling feature/well type	There are no controlled vocabularies within the current IGSN template to characterize the type of well. We currently recommend providing this information in the free-text location description.

Table 3. Summary of preliminary issues and solutions encountered in assigning SESAR IGSN metadata to sample locations. While the most basic location information is included (e.g. latitude, longitude, and location description), more work is needed on interoperability with standards that more fully describe site locations, such as metadata standards developed by the Open Geospatial Consortium. Location descriptions in multidisciplinary ecosystem sciences include location descriptions for samples and other entities, such as sensor infrastructure in monitoring networks and remote sensing data.